

Bodyfat Percentage  
(Regression)

The goal of this study is to train a model in order to predict the percentage of bodyfat in men. The dataset used in this case study is found in <https://www.kaggle.com/datasets/fedesoriano/body-fat-prediction-dataset> and has 14 features and 251 samples. This dataset contains information on certain factors like Density determined from underwater weighing, age, weight, height, neck circumference, Chest circumference, Abdomen 2 circumference, Hip circumference, Thigh circumference, Knee circumference, Ankle circumference, Biceps (extended) circumference, Forearm circumference, Wrist circumference. The dataset contains no missing values and also all features are numerical, therefore there is no need to for transforming any of them.

# Step 1: Import data from file

Right click on the input spreadsheet and choose the option “Import from file”. Then navigate through your files to load the one with the housing price data.

A screenshot of a spreadsheet

Description automatically generatedA screenshot of a computer

Description automatically generated

# Step 2: Manipulate data

In order to use the data for training we have to exclude any columns that do not contain features. In our dataset there are no such columns. Therefore, we will include all columns in the training. We follow these steps to execute this:

* On the menu click on “Data Transformation” → “Data Manipulation” → “Select Column(s)”
* Select all columns.

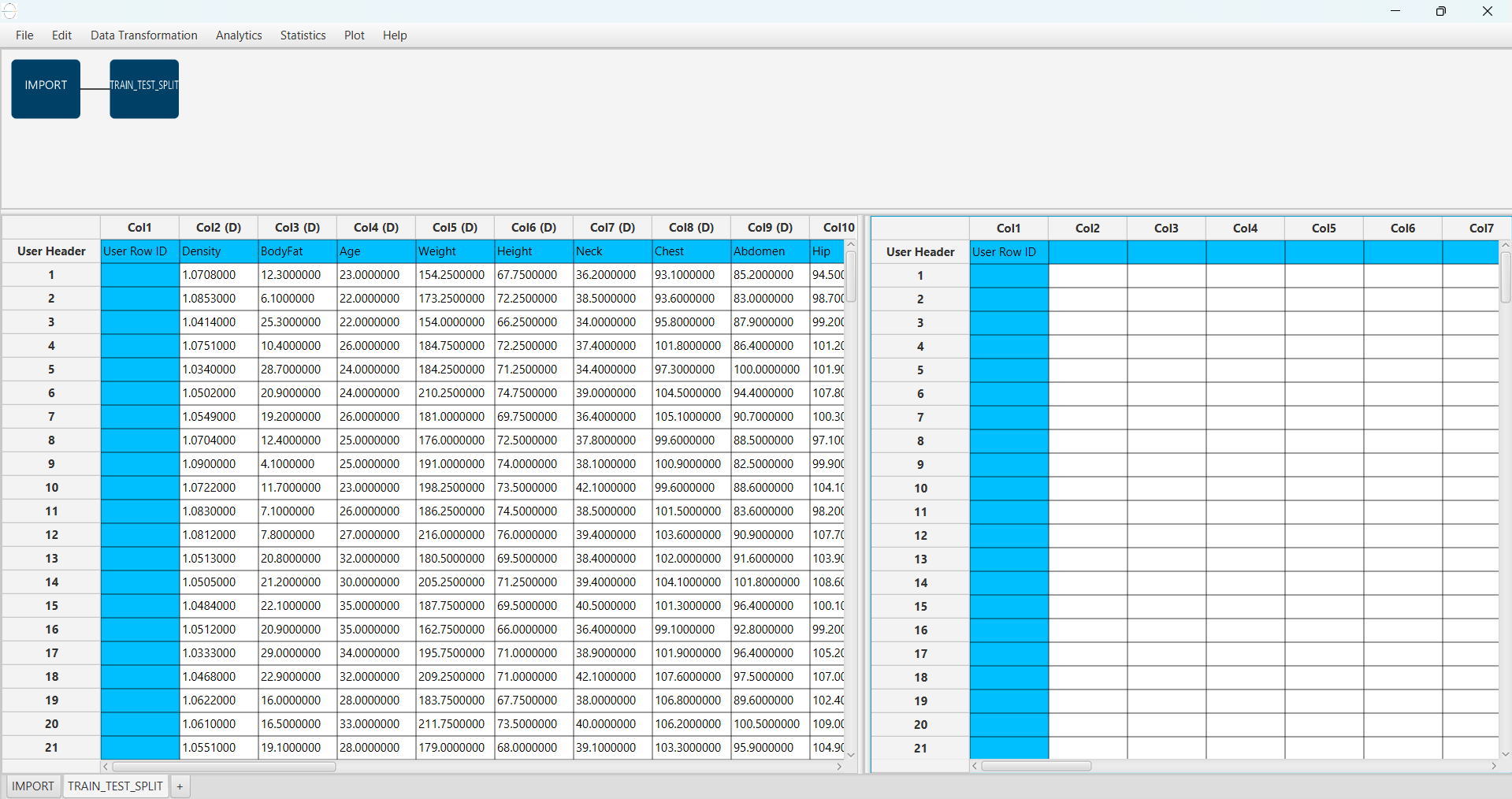
|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated | A screenshot of a computer  Description automatically generated |

The data will appear in the output spreadsheet.

# Step 3: Split data

Create a new tab by pressing the “+” button on the bottom of the page with the name “TRAIN\_TEST\_SPLIT” which we will use for splitting to create the train and test set.

Import data into the input spreadsheet of the “TRAIN\_TEST\_SPLIT” tab from the output of the “IMPORT” tab by right-clicking on the input spreadsheet and then choosing “Import from SpreadSheet”.



Split the dataset by choosing from the top ribbon: “Data Transformation” → “Split” → “Random Partitioning”. Then choose the “Training set percentage” and the column for the sampling as shown below:

|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated | A screenshot of a computer  Description automatically generated |

The results will appear on the output spreadsheet.

# A screenshot of a computer Description automatically generatedStep 4: Normalize the training set

Create a new tab by pressing the “+” button on the bottom of the page with the name “NORMALIZE\_TRAIN\_SET”.

Import data into the input spreadsheet of the “NORMALIZE\_TRAIN\_SET” tab the train set from the output of the “TRAIN\_TEST\_SPLIT” tab by right-clicking on the input spreadsheet and then choosing “Import from SpreadSheet”. From the available Select input tab options choose “TRAIN\_TEST\_SPLIT: Training Set”.

A screenshot of a computer

Description automatically generated

Normalize the data using Z-score by browsing: “Data Transformation” → “Normalizers” → “Z-Score”. Then select all columns except “BodyFat” and click “Execute”.

|  |  |
| --- | --- |
|  | A screenshot of a computer  Description automatically generated |

The results will appear on the output spreadsheet.

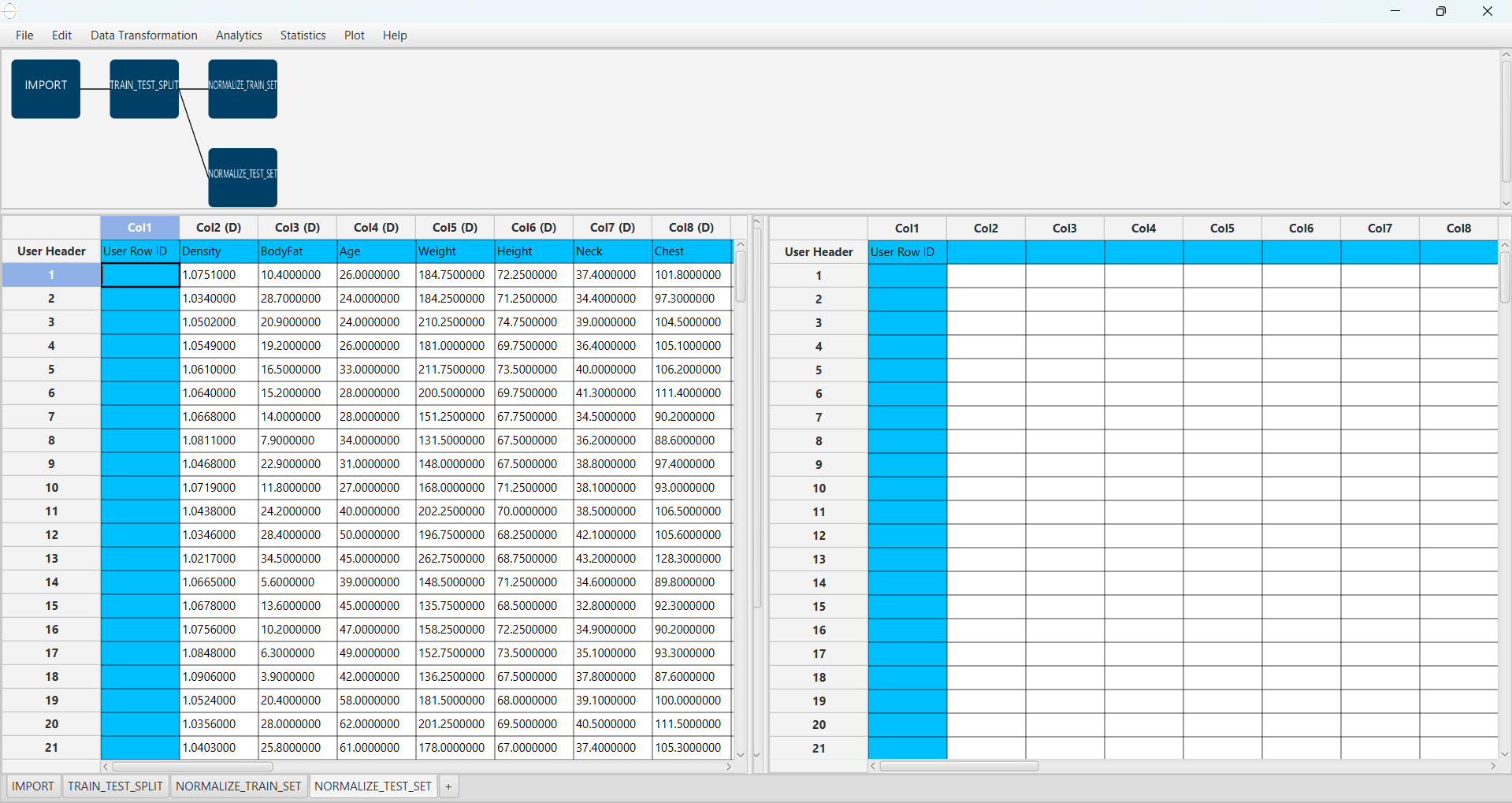
A screenshot of a computer

Description automatically generated

# Step 5: Normalize the test set

Create a new tab by pressing the “+” button on the bottom of the page with the name “NORMALIZE\_TEST\_SET”.

Import data into the input spreadsheet of the “NORMALIZE\_TEST\_SET” tab the test set from the output of the “TRAIN\_TEST\_SPLIT” tab by right-clicking on the input spreadsheet and then choosing “Import from SpreadSheet”. From the available Select input tab options choose “TRAIN\_TEST\_SPLIT: Test Set”.



Normalize the test set using the existing normalizer of the training set by browsing:   
“Analytics” → “Existing Model Utilization” → “Model (from Tab:) NORMALIZE\_TRAIN\_SET”.

|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated |  |

The results will appear on the output spreadsheet.

A screenshot of a computer

Description automatically generated

# Step 6: Train the model

Create a new tab by pressing the “+” button on the bottom of the page with the name “TRAIN\_MODEL(.fit)”. Import data into the input spreadsheet of the “TRAIN\_MODEL(.fit)” tab from the output of the “NORMALIZE\_TRAIN\_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from SpreadSheet”.

A screenshot of a computer

Description automatically generated

Use the k Nearest Neighbors (kNN) method to train and fit the model by browsing:   
“Analytics” → “Regression” → “k Nearest Neighbors (kNN)” and set the “Target Column” as the column corresponding to “BodyFat” and the “Number of Neighbors” to 5.

|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated | A screenshot of a computer  Description automatically generated |

The predictions will appear on the output spreadsheet.

A screenshot of a computer

Description automatically generated

# Step 7: Validate the model

Create a new tab by pressing the “+” button on the bottom of the page with the name “VALIDATE\_MODEL(.predict)”.

Import data into the input spreadsheet of the “VALIDATE\_MODEL(.predict)” tab from the output of the “NORMALIZE \_TEST\_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from SpreadSheet”.

A screenshot of a computer

Description automatically generated

To validate the model browse: “Analytics” → “Existing Model Utilization”. Then choose Model “(from Tab:) TRAIN\_MODEL (.fit)”. and transfer the “BodyFat” column in the output.

|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated | A screenshot of a computer  Description automatically generated |

The predictions will appear on the output spreadsheet.

A screenshot of a computer

Description automatically generated

# Step 8: Statistics calculation

Create a new tab by pressing the “+” button on the bottom of the page with the name “STATISTICS\_ACCURACIES”.

Import data into the input spreadsheet of the “STATISTICS\_ACCURACIES” tab from the output of the “VALIDATE\_MODEL(.predict)” tab by right-clicking on the input spreadsheet and then choosing “Import from SpreadSheet”.

A screenshot of a computer

Description automatically generated

Calculate the statistical metrics for the regression by browsing: “Statistics” → “Model Metrics” → “Regression Metrics”.

|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated | A screenshot of a computer  Description automatically generated |

The results will appear on the output spreadsheet.

A screenshot of a computer

Description automatically generated

# Step 9: Reliability check of each record of the test set

## Step 9.a: Create the domain

Create a new tab by pressing the “+” button on the bottom of the page with the name “EXCLUDE\_BODYFAT”.

Import data into the input spreadsheet of the “EXCLUDE\_ BODYFAT” tab from the output of the “NORMALIZE \_TRAIN\_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from SpreadSheet”.

A screenshot of a computer

Description automatically generated

Manipulate the data to exclude the column that corresponds to the “BodyFat” by browsing: “Data Transformation” → “Data Manipulation” → “Select Columns”. Then select all the columns except the “BodyFat”. The filtered table will appear on the output spreadsheet.

|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated | A screenshot of a computer  Description automatically generated |

Create a new tab by pressing the “+” button on the bottom of the page with the name “DOMAIN”.

Import data into the input spreadsheet of the “DOMAIN” tab from the output of the “EXCLUDE\_ BODYFAT” tab by right-clicking on the input spreadsheet and then choosing “Import from SpreadSheet”.

A screenshot of a computer

Description automatically generated

Create the domain of applicability by browsing: “Statistics” → “Domain APD”.

|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated | A screenshot of a computer  Description automatically generated |

The results will appear on the output spreadsheet.

A screenshot of a computer

Description automatically generated

## Step 9.b: Check the test set reliability

Create a new tab by pressing the “+” button on the bottom of the page with the name “EXCLUDE\_BODYFAT\_TEST\_SET”.

Import data into the input spreadsheet of the “EXCLUDE\_ BODYFAT \_TEST\_SET” tab from the output of the “NORMALIZE \_TEST\_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from SpreadSheet”.

A screenshot of a computer

Description automatically generated

Filter the data to exclude the column that corresponds to the “BodyFat” by browsing: “Data Transformation” → “Data Manipulation” → “Select Columns”. Then select all the columns except “BodyFat”.

|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated | A screenshot of a computer  Description automatically generated |

The results will appear on the output spreadsheet.

Create a new tab by pressing the “+” button on the bottom of the page with the name “RELIABILITY”.

Import data into the input spreadsheet of the “RELIABILITY” tab from the output of the “EXCLUDE\_ BODYFAT \_TEST\_SET” tab by right-clicking on the input spreadsheet and then choosing “Import from SpreadSheet”.

A screenshot of a computer

Description automatically generated

Check the predictions’ reliability by browsing: “Analytics” → “Existing Model Utilization”. Then select as Model “(from Tab:) DOMAIN”.

|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated | A screenshot of a computer  Description automatically generated |

The results will appear on the output spreadsheet. We can observe that there are no unreliable samples in the test set.

A screenshot of a computer

Description automatically generated

# Final Isalos Workflow

Following the above-described steps, the final workflow on Isalos will look like this:

A screenshot of a computer

Description automatically generated